



THE UNIVERSITY *of* EDINBURGH

## Edinburgh Research Explorer

# A MILI-independent piRNA biogenesis pathway empowers partial germline reprogramming

### Citation for published version:

Vasilauskaite, L, Vitsios, D, Berrens, RV, Carrieri, C, Reik, W, Enright, AJ & O'Carroll, D 2017, 'A MILI-independent piRNA biogenesis pathway empowers partial germline reprogramming', *Nature Structural & Molecular Biology*, vol. 24, no. 7, pp. 604–606. <https://doi.org/10.1038/nsmb.3413>

### Digital Object Identifier (DOI):

[10.1038/nsmb.3413](https://doi.org/10.1038/nsmb.3413)

### Link:

[Link to publication record in Edinburgh Research Explorer](#)

### Document Version:

Peer reviewed version

### Published In:

Nature Structural & Molecular Biology

### General rights

Copyright for the publications made accessible via the Edinburgh Research Explorer is retained by the author(s) and / or other copyright owners and it is a condition of accessing these publications that users recognise and abide by the legal requirements associated with these rights.

### Take down policy

The University of Edinburgh has made every reasonable effort to ensure that Edinburgh Research Explorer content complies with UK legislation. If you believe that the public display of this file breaches copyright please contact [openaccess@ed.ac.uk](mailto:openaccess@ed.ac.uk) providing details, and we will remove access to the work immediately and investigate your claim.



# **A MILI-independent piRNA biogenesis pathway empowers partial germline reprogramming.**

Lina Vasilauskaitė<sup>1,2</sup>, Dimitrios Vitsios<sup>3</sup>, Rebecca Berrens<sup>4</sup>, Claudia Carrieri<sup>1,2</sup>, Wolf Reik<sup>4,5</sup>, Anton J Enright<sup>3</sup> and Dónal O’Carroll<sup>1,2</sup>

1. European Molecular Biology Laboratory (EMBL), Monterotondo, Italy.
2. MRC Centre for Regenerative Medicine, School of Biological Sciences, University of Edinburgh, Edinburgh, UK.
3. European Bioinformatics Institute, Hinxton, Cambridge, UK.
4. Epigenetics Programme, Babraham Institute, Cambridge, UK.
5. Wellcome Trust Sanger Institute, Hinxton, Cambridge, UK.

Corresponding author: Dónal O’Carroll

Email: donal.ocarroll@ed.ac.uk

**In mice, the PIWI-piRNA pathway is essential to re-establish transposon silencing during male germline reprogramming. The cytoplasmic PIWI protein MILI mediates piRNA-guided transposon RNA cleavage as well as piRNA amplification. MIWI2-bound piRNAs and its nuclear localization are proposed to be dependent upon MILI function. Here, we demonstrate the existence of a piRNA biogenesis pathway that in the absence of MILI sustains partial MIWI2 function and reprogramming activity.**

The erasure of DNA methylation during male mammalian germline reprogramming and the concomitant de-repression of transposons initiate the process of re-silencing through the production of primary piRNAs<sup>1</sup>. In mice, these piRNAs are bound to the cytoplasmic PIWI protein MILI, which is a piRNA-directed RNA endonuclease that mediates piRNA amplification of secondary piRNAs to ensure transposon repression<sup>2</sup> (Supplementary Fig. 1). These secondary piRNAs through sequence complementarity guide the PIWI protein MIWI2 to direct *de novo* DNA methylation of LINE1 and IAP elements<sup>1-5</sup> (Supplementary Fig. 1). The loss of MILI or MIWI2 results in defective *de novo* DNA methylation and depression of LINE1s and IAPs causing meiotic arrest<sup>3,5,6</sup>. In addition to meiotic arrest, *Miwi2*<sup>-/-</sup> mice progressively lose germ cells and are fully aspermatogenic by 9 months<sup>2,3</sup>. We found that *Mili*<sup>-/-</sup> mice present a milder germ cell loss phenotype with ~50% tubules being spermatogenic at 1 year of age (Fig. 1a,b). This discrepancy in phenotype indicates that MIWI2 may have a MILI-independent role during fetal germ cell reprogramming. We therefore FACS-purified undifferentiated spermatogonia and performed whole genome bisulfite sequencing. While the loss of MILI and MIWI2 predominantly affects LINE1 and IAP methylation<sup>4,5,7</sup> (Supplementary Fig. 2), we found in accordance with our hypothesis that *Miwi2*-deficiency has a more severe impact on genome re-methylation than *Mili*-disruption (Fig. 1c and Supplementary Fig. 2). This contrasts with a previous report<sup>8</sup>; the discrepancy likely arises from the fact that in the previous study *Mili*<sup>-/-</sup> and *Miwi2*<sup>-/-</sup> spermatogonia were differentially isolated with varying degrees of somatic cell contamination<sup>8</sup>. These somatic cells have normal genomic methylation levels and thus a differential degree of contamination confounds a quantitative comparative analysis. Here we observed that the methylation of 1704 loci were dependent upon MIWI2 function, whereas 258 loci showed selective dependence upon MILI (Fig. 1c). To determine if the methylation of these loci could be attributed to the piRNA pathway, we analyzed whether piRNAs bound to MILI or MIWI2 fetal gonad RNPs map to these loci. Indeed, piRNAs from both RNPs were complimentary to these loci (Fig. 1d). We next analyzed uniquely mapping piRNAs, which would exclude transposon-associated

piRNAs and found that unique piRNAs from the respective RNPs map to both MIWI2- and MILI-dependent loci (Fig. 1e). In summary, *Miwi2*-deficiency has a greater impact on spermatogenesis and fetal reprogramming in comparison to the loss of *Mili*.

The physiological and molecular discrepancy between *Miwi2*- and *Mili*-deficiency could indicate that a non-canonical piRNA biogenesis and partial piRNA-dependent MIWI2 function exists in the absence of MILI. To test this hypothesis, we generated total small RNA libraries from wild type and *Mili*<sup>-/-</sup> E16.5 fetal testis. Albeit at low levels, a clear presence of piRNAs is found in the *Mili*<sup>-/-</sup> libraries (Fig. 2a and Supplementary Fig 3a). Importantly, these are *bona fide* piRNAs as all the piRNAs found in absence of MILI are present in wild type fetal gonads (Fig. 2b). LINE1 and IAP piRNAs are generated in greatly reduced quantities and as expected the hallmarks of piRNA amplification are absent in the *Mili*<sup>-/-</sup> libraries (Supplementary Fig. 3b-e). In the *Mili*<sup>-/-</sup> mutant, piRNAs cognate to the 1704 MIWI2-dependent loci were still present but at lower levels (Fig. 2c and Supplementary Fig. 3f), with a 2- and 3-fold reduction in multi or unique-mapping piRNAs observed, respectively (Supplementary Fig. 3g). In contrast, in the *Mili*<sup>-/-</sup> libraries there was a drastic 18-fold reduction in the uniquely mapping piRNAs to the 256 loci whose methylation is dependent upon MILI function (Supplementary Fig. 3g). Since piRNAs exist in *Mili*<sup>-/-</sup> fetal gonocytes albeit in reduced quantities, this should result in the partial nuclear localization of MIWI2. We next analyzed the localization of MIWI2 in E16.5 *Mili*<sup>-/-</sup> fetal gonocytes by confocal immunofluorescence using a rabbit polyclonal antibody (Fig. 2d and Supplementary Fig. 4a). This analysis revealed that a portion of MIWI2 indeed retains nuclear localization (Fig. 2d). Quantification of the signal in wild type and *Miwi2*<sup>-/-</sup> fetal testis revealed that the anti-MIWI2 antibody gave a non-negligible level of non-specific background staining (Supplementary Fig. 4b). To independently confirm this finding, we generated an N-terminal HA epitope tagged MIWI2 (HA-MIWI2) mouse allele (*Miwi2*<sup>HA</sup>) (Supplementary Fig. 5), permitting the use of high quality monoclonal HA antibodies. Confocal immunofluorescence of HA-MIWI2 in *Mili*<sup>-/-</sup> fetal gonocytes confirmed the partial MIWI2 nuclear localization with dramatically reduced levels of non-specific background (Fig. 2e and Supplementary Fig. 4c,d).

The selective loss of MILI's endonuclease activity attenuates piRNA amplification, but leaves primary piRNA biogenesis intact, resulting in a MIWI2 RNP that can partially function to silence IAP<sup>2</sup>, whereas the complete loss of MILI function was thought to abrogate all piRNA biogenesis and MIWI2 reprogramming activity<sup>1</sup>. Here, we show that a non-canonical primary



piRNA biogenesis exists in the absence of MILI which results in partial MIWI2 nuclear localization and reprogramming activity. This residual reprogramming activity is supported by the fact that more severe methylation and spermatogenesis defects are observed in *Miwi2*<sup>-/-</sup> over *Mili*<sup>-/-</sup> mice. The piRNA biogenesis in *Mili*<sup>-/-</sup> mice may resemble the phased primary piRNA pathway identified in *Drosophila*, *Bombyx mori* and mouse, where the binding of PIWI to cleaved transcripts from the amplification cycle coupled with Zucchini-mediated endonucleolytic cleavage results in additional piRNAs biogenesis<sup>9-12</sup>. While phased piRNA biogenesis in *Drosophila* can occur in the absence of piRNA amplification<sup>13</sup>, we could not detect phasing in the *Mili*<sup>-/-</sup> piRNA population (data not shown), likely given the extensive 3' trimming that occurs during mouse piRNA biogenesis<sup>14</sup>. In summary, here we show that mouse fetal primary piRNA biogenesis can exist in the absence of MILI and this residual piRNA pathway contributes to germline reprogramming.

## METHODS

Methods and any associated references are available in the online version of the paper.

**Accession codes.** Small RNA sequencing data that support the findings of this study have been deposited in [www.ebi.ac.uk/ena](http://www.ebi.ac.uk/ena) with the accession code PRJEB19792. WGBS data that support the findings of this study have been deposited in [www.ebi.ac.uk/arrayexpress](http://www.ebi.ac.uk/arrayexpress) with the accession code E-MTAB-5561. MIWI2 and MILI RNP-associated piRNA libraries were used from previously published study<sup>2</sup>.

*Note: Any Supplementary Information and Source Data files are available in the online version of the paper. The imaging data that support the findings of this study are available from the corresponding author upon reasonable request.*

## ACKNOWLEDGEMENTS

The research leading to these results has received funding from the European Research Council under the European Union's Seventh Framework Programme (FP7/2007-2013) / ERC grant agreement n° GA 310206 awarded to DO'C.

## AUTHOR CONTRIBUTIONS

L.V. contributed to the design, execution and analysis of the majority of experiments. D.V. performed the bioinformatic piRNA analysis. R.B. generated the whole genome bisulfite

libraries and performed the bioinformatic analysis of genomic methylation. C.C. designed, generated and validated the *Miwi2*<sup>HA</sup> allele. A.J.E. and W.R. oversaw all bioinformatic analyses performed. D.O'C. conceived and supervised this study. L.V. and D.O'C wrote the final version of the manuscript.

## COMPETING FINANCIAL INTERESTS

The authors declare no competing financial interests.

1. Aravin, A.A. et al. *Mol Cell* **31**, 785-99 (2008).
2. De Fazio, S. et al. *Nature* **480**, 259-63 (2011).
3. Carmell, M.A. et al. *Dev Cell* **12**, 503-14 (2007).
4. Kuramochi-Miyagawa, S. et al. *Genes Dev* **22**, 908-17 (2008).
5. Aravin, A.A., Sachidanandam, R., Girard, A., Fejes-Toth, K. & Hannon, G.J. *Science* **316**, 744-7 (2007).
6. Kuramochi-Miyagawa, S. et al. *Development* **131**, 839-49 (2004).
7. Molaro, A. et al. *Genes Dev* **28**, 1544-9 (2014).
8. Manakov, S.A. et al. *Cell Rep* **12**, 1234-43 (2015).
9. Han, B.W., Wang, W., Li, C., Weng, Z. & Zamore, P.D. *Science* **348**, 817-21 (2015).
10. Mohn, F., Handler, D. & Brennecke, J. *Science* **348**, 812-7 (2015).
11. Homolka, D. et al. *Cell Rep* **12**, 418-28 (2015).
12. Yang, Z. et al. *Mol Cell* **61**, 138-52 (2016).
13. Wang, W. et al. *Mol Cell* **59**, 819-30 (2015).
14. Saxe, J.P., Chen, M., Zhao, H. & Lin, H. *EMBO J* **32**, 1869-85 (2013).

## FIGURE LEGENDS

**Figure 1 Loss of MIWI2 results in more severe physiological and molecular phenotype than MILI-deficiency. (a)** Percentage of aspermatogenic tubules in *Mili*<sup>+/+</sup> and *Mili*<sup>-/-</sup> 1-year-old mice. Error bars represent a standard deviation of the mean (n=5 animals). **(b)** Representative images of hematoxylin and eosin stained testis cross-sections of *Mili*<sup>+/+</sup> and *Mili*<sup>-/-</sup> 1-year-old mice. **(c)** Comparisons of the percentage of CpG methylation in wild type (WT) and *Mili*<sup>-/-</sup>, WT and *Miwi2*<sup>-/-</sup> as well as *Mili*<sup>-/-</sup> and *Miwi2*<sup>-/-</sup> undifferentiated spermatogonia are shown. Blue dots represent significantly differentially methylated regions (DMRs). **(d & e)** piRNAs (d) and uniquely mapping piRNAs (e) from MIWI2 and MILI ribonucleoprotein complexes (RNPs) mapped to loci whose methylation is dependent upon

MIWI2 and MILI, respectively. Positive (red) and negative (blue) values indicate sense and antisense piRNAs, respectively. An averaged value from biological duplicates is shown (n=2 animals).

**Figure 2 piRNAs are produced and MIWI2 is partially localized to the nucleus in the absence of MILI.** (a) Length distribution of piRNAs in wild type (WT) and *Mili*<sup>-/-</sup> E16.5 fetal testis. (b) piRNA expression analysis in WT and *Mili*<sup>-/-</sup> E16.5 fetal testis. (c) Unique piRNAs from WT and *Mili*<sup>-/-</sup> E16.5 fetal testis mapped to loci whose methylation is dependent upon MIWI2 and MILI, respectively. Positive (red) and negative (blue) values indicated sense and antisense piRNAs, respectively. An averaged value from biological duplicates is shown in panels a-c (n=2 animals). (d & e) Immunofluorescence with anti-MIWI2 antibody (d) and anti-HA antibody (e) on fetal testis of the indicated genotypes with a single fetal gonocyte shown. Scan-line profile plots (bottom panel) represent relative fluorescence intensity of MIWI2 (d) and HA (e) (green) staining in the nucleus (blue) along the analyzed trajectory (represented as arrow in d and e).

## ONLINE METHODS

**Mouse strains and experimentation.** The *Mili*<sup>-</sup> and *Miwi2*<sup>-</sup> mouse alleles were previously produced in the O'Carroll laboratory<sup>2,15</sup> and maintained on a mixed 129/C57Bl/6 genetic background. The *Miwi2-tdTomato* transcriptional allele was used to isolate spermatogonia (Carrieri C and O'Carroll D, unpublished observations). The *Miwi2*<sup>HA</sup> allele was created using CRISPR/Cas9 gene editing technology<sup>16</sup>. The sequence encoding a HA-tag was inserted after the starting ATG codon of *Miwi2* to generate the *Miwi2*<sup>HA</sup> allele that results in an N-terminal HA-MIWI2 fusion protein. An sgRNA (5' ACCCAGGAACATGAGTGGAC) targeting *Miwi2* in vicinity of the starting ATG codon was used. A DNA oligonucleotide with HA and BamHI restriction site flanked by 60bp of 3' and 5' *Miwi2* genomic sequence was used as a template for homology directed repair. The sgRNA (12,5 ng/ul), DNA oligonucleotide (12,5 ng/ul), as well as Cas9 mRNA (25 ng/ul) were injected into the cytoplasm of two pronuclei stage C57Bl/6 mouse embryos. Manipulated embryos were cultured overnight and embryos that had proceeded to the 2 cell stage were transferred into the oviduct of pseudopregnant females. The presence of *Miwi2*<sup>HA</sup> allele in mice was screened by digesting PCR products amplified by primers (Fw 5'ACAGCCACACCGTCTCTTTT; Rv 5'CAGGATAGCCAAAGGAAGGA) flanking *Miwi2* ATG with BamHI. The *Miwi2*<sup>HA</sup> allele was also confirmed by sequencing. For the analysis of progressive germ cell loss, *Mili*<sup>+/-</sup> and

*Milt*<sup>-/-</sup> mice were aged to one year, sacrificed and testes collected. Each genotype group contained five animals to ensure adequate power to detect and describe differences between the genotypes. E16.5 fetal testes were used for immunofluorescence analysis and small RNA library preparation. No blind experiments or randomization of samples or animals were employed in this study. All mice used in this study were male; and bred and maintained in EMBL Mouse Biology Unit, Monterotondo, and subsequently in the Centre for Regenerative Medicine, Edinburgh. All procedures were done in accordance to the current Italian legislation (Art. 9, 27. Jan 1992, nu116) under license from the Italian health ministry or the UK Home Office regulations, respectively.

**Spermatogonia isolation.** Isolated postnatal day 14 testes from mice of the respective genotypes that also carried the *Miwi2-tdTomato* transcriptional allele in heterozygosity were dealbulgated and digested with collagenase (0,5 mg/ml, Sigma) at 32°C for 10 minutes shaking in 1ml of DMEM media (Life technologies) supplemented with Pen/Strep (Life technologies), NEAA (Life technologies), sodium pyruvate (Life technologies) and sodium lactate (Sigma). After digestion seminiferous tubules were collected by centrifugation at 1000 rpm for 5 min. Single cell suspension was prepared by further digesting seminiferous tubules in 1 ml of 0,05% trypsin (Life technologies) and DNase (Sigma), added to 0,05 mg/ml of final concentration, at 32°C for 8-10 minutes shaking. After digestion trypsin was neutralized by adding 100 µl of FCS. Cells were spun and resuspended in PBS with 3% FCS. Germ cells were stained with following antibodies: c-Kit 1:800 (eBioscience 25-1171), CD45-biotin 1:400 (eBioscience 13-0451), CD51-biotin 1:100 (Biolegend 104104), streptavidin-qDot 1:50 (eBioscience 93-4317), CD9 1:50 (eBioscience 11-009) in PBS with 3% of FCS and 0,01% sodium azide. SYTOX Blue (Life technologies) was used as a living dye. Target population was selected by excluding cell doublets and CD45<sup>pos</sup> and CD51<sup>pos</sup> cells by initially choosing to analyze cells with low side scatter and negative for live cell dye SYTOX Blue. Among CD45<sup>neg</sup> and CD51<sup>neg</sup> cells, a target population was identified as c-Kit<sup>neg</sup> and having a live fluorescence of tdTomato and being CD9<sup>pos</sup>. Cells were sorted with an 85µm nozzle using a FACSaria II SORP (BD Biosciences). This strategy results in the isolation of undifferentiated spermatogonia with a purity of greater than 98%.

**Histology.** Testes were fixed in Bouin's fixative (Sigma) overnight at 4°C and paraffin embedded. 8 µm thick sections were stained with hematoxylin and eosin using routine methods.

**Immunofluorescence.** Testes were freshly embedded into OCT compound (Sakura), 8  $\mu$ m sections were cut, fixed in 4% paraformaldehyde (Sigma) 10 minutes at room temperature and then permeabilized for 10 minutes at room temperature in 0.1% Triton-X100 (Sigma). Subsequently sections were blocked for 30 minutes at room temperature in 10% normal donkey serum (Sigma), 1% BSA (Sigma) and 0.1M glycine (Sigma). Primary antibody incubation was done overnight at 4°C in the blocking buffer. The following antibodies were used at indicated dilutions: rabbit polyclonal anti-MIWI2 1:200<sup>2</sup>, rabbit monoclonal anti-HA 1:500 (Cell Signaling 3724). Appropriate donkey anti-rabbit AlexaFluor488 secondary antibody (Life technologies) was used in a dilution 1:1000. DAPI (5  $\mu$ g/ $\mu$ l) (Life technologies) was used to stain DNA. Samples were mounted with ProLong Gold antifade reagent (Life technologies). Leica TCS SP5 confocal microscope was used to acquire images. Five images were acquired per each biological replicate per genotype. After acquisition images were processed with ImageJ and Adobe Photoshop CS5 computer programs. All images corresponding to the same experiment were acquired and processed applying the same settings. To determine MIWI2 and HA immunofluorescence intensity in the nuclei of gonocytes, fetal testes cross-sections were stained with MIWI2 and HA antibodies as described above. Fetal testes from two embryos were used per each genotype analyzed as biological duplicates, except a biological triplicate was used for *Miwi2*<sup>-/-</sup> fetal testis. All the data was collected in 2-3 independent immunofluorescence experiments. Acquired stacks of images were merged using ImageJ computer program. ROI (region of interest) manager tool in ImageJ was used to quantify relative fluorescence intensity in the nuclei, where ROIs were manually defined, avoiding cytoplasmic signal.

**Bisulfite conversion.** For bisulfite conversion DNA was isolated from sorted cells as described above. Whole genome bisulfite sequencing libraries were generated using a post-bisulfite adaptor tagging (PBAT) method as previously described<sup>17</sup> using 10 cycles of PCR amplification. Libraries were sequenced using Illumina HiSeq 2000. Three biological replicates (cells independently isolated from three animals) were generated per genotype to ensure adequate power to detect statistically significant differences between the respective genotypes.

**Bisulfite sequencing analysis.** Raw sequence reads were trimmed to remove both poor quality calls and adapters using Trim Galore (v0.4.1,

[www.bioinformatics.babraham.ac.uk/projects/trim\\_galore/](http://www.bioinformatics.babraham.ac.uk/projects/trim_galore/), Cutadapt version 1.8.1, parameters: --paired). Trimmed reads were aligned to the mouse genome (GRCm38) in paired-end mode to be able to use overlapping parts of the reads only once. Alignments were carried out with Bismark v0.14.4<sup>18</sup> with the following set of parameters: paired-end mode: --pbat. Reads were then deduplicated with deduplicate\_bismark selecting a random alignment for position that was covered more than once. CpG methylation calls were extracted from the deduplicated mapping output ignoring the first 6 bp of each read to reduce the methylation bias typically observed in PBAT libraries using the Bismark methylation extractor (v0.14.4) with the following parameters: a) paired-end mode: --ignore 6 --ignore\_r2 6; b) single-end mode: --ignore 6. CpG methylation calls were analysed using R and SeqMonk software ([www.bioinformatics.babraham.ac.uk/projects/seqmonk/](http://www.bioinformatics.babraham.ac.uk/projects/seqmonk/)). 50 adjacent CpG running window probes were generated and percentage of methylation determined for probes containing at least 5 reads and 3 CpG on the pooled replicate data. The methylation level was expressed as the mean of individual CpG sites. Graphing and statistics were performed using Seqmonk and RStudio.

**Transposon analysis in bisulfite sequencing data.** Repeat locations for a pre-defined set of repeat classes of interest were extracted from the pre-masked repeatmasker libraries (mm10 - Dec 2011 - RepeatMasker open-4.0.5, <http://www.repeatmasker.org/species/musMus.html>). Repeat instances overlapping annotated genes in the Ensembl gene set were removed to avoid mixing signals from genic expression with specific expression of repetitive sequences. Methylation levels at the repeat instances were quantitated by summing up all methylation calls and non-methylation calls for all instances of each class of repeat and calculating the percentage of methylated Cs over all Cs.

**Small RNA libraries.** Isolated E16.5 fetal testes were homogenized in 1 ml of Qiazol (QIAGEN). Extracted total fetal testes RNA was used to generate small RNA libraries using NEBNext Multiplex Small RNA Library Prep Set for Illumina (Set 1) (NEB) following manufacturer's instructions with 20 cycles of PCR amplification. Small RNA libraries were generated from two biological replicates (RNA isolated from fetal testes from two animals) per genotype to ensure adequate confidence in observations.

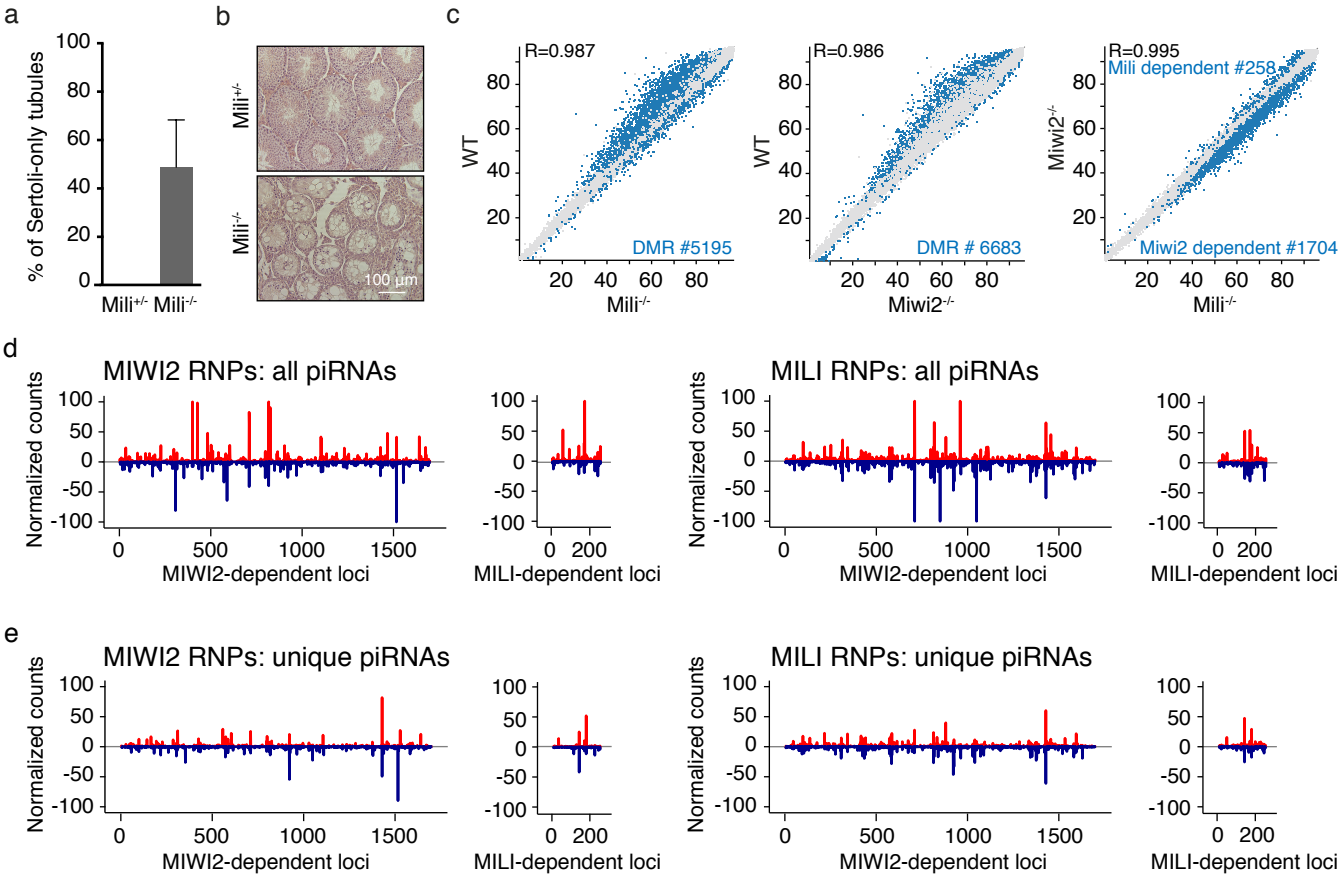
**Small RNA sequencing analysis.** All samples were initially aligned against Rfam<sup>19</sup> in order to filter out tRNA sequences (tRNA hits with an alignment identity score > 90% were

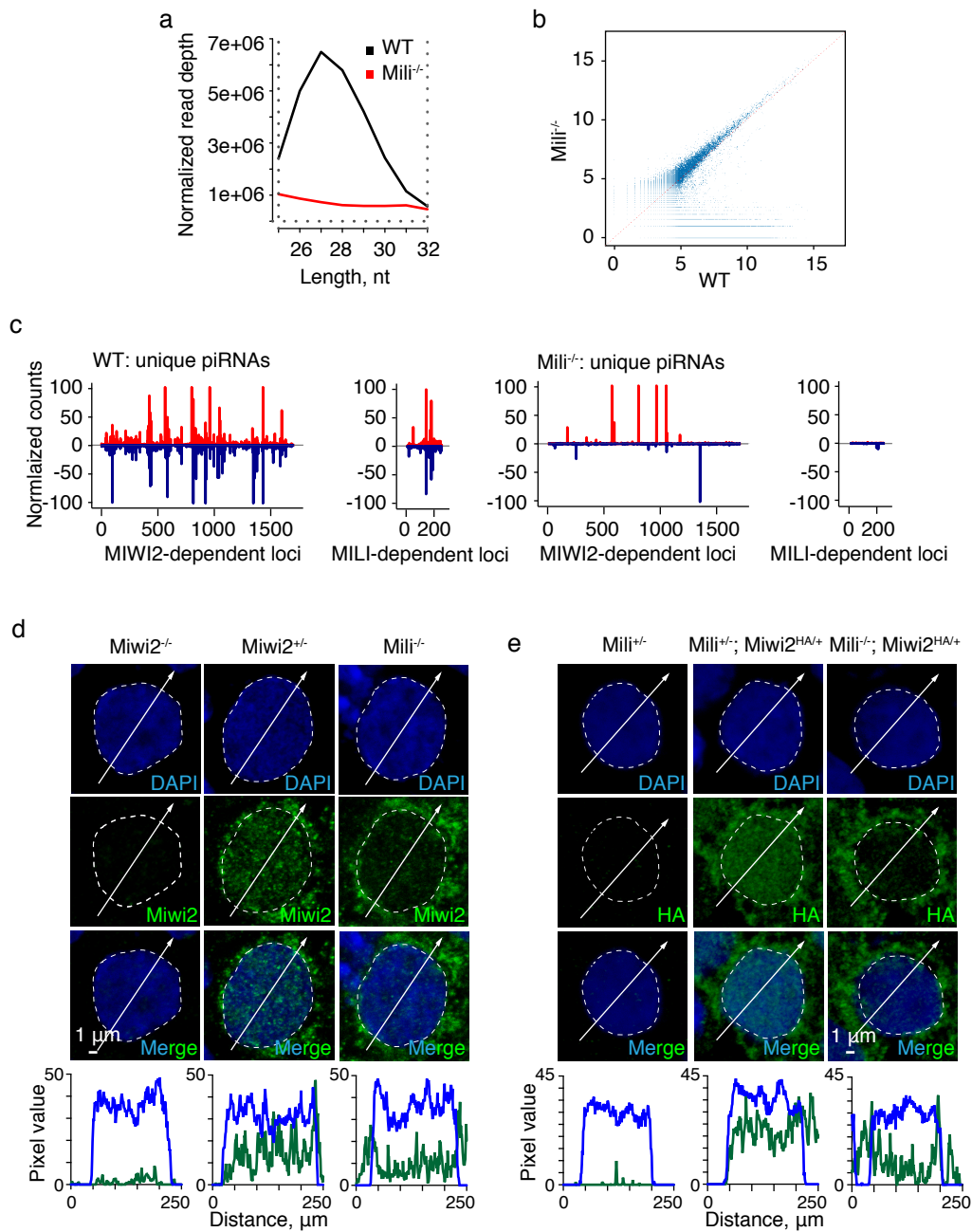
excluded from the rest of the analysis). Analysis of the filtered samples was then performed using *SequenceImp*<sup>20</sup>. Input reads were first trimmed from the 3' adapter with *reaper* (using default configuration for read geometry without barcode) and de-duplicated with *tally*, which are both part of the *Kraken* suite of tools<sup>20</sup>. The length distribution of all cleaned reads between 18 and 32 nucleotides was recorded in order to check for depletion of piRNA sequences between the wild type and *Mili*<sup>-/-</sup> conditions. Cleaned reads were later filtered by length (retaining only 24-32 nucleotide long sequences) and aligned against the Mouse genome (Ensembl release 66) allowing up to 2 mismatches and reporting up to 20 hits per sequence, when analyzing for all mappers. In the case of uniquely mapped sequences, *bowtie* call from the *SequenceImp* pipeline was tweaked using the parameter *-m* = 1 (parameter *-k* was set to 1 for both the unique and all mappers cases). BAM output files from the alignment step were intersected with BED files containing the coordinates of 1704 and 258 loci, whose methylation is dependent on MIWI2 and MILI, respectively. PiRNA counts within each locus were calculated as the average number of fragments aligning against the locus, divided by the size of the locus region in 1 kb units. Expression densities were limited within the interval (-100, 100) in order to filter out outliers from 3 overexpressed loci and thus increase densities resolution for all loci. With regards to the piRNA differential expression analysis, a custom database of all 26-31 nucleotide long unique sequences found across all wild type and *Mili*<sup>-/-</sup> replicates was initially built. Each sample replicate was then aligned against this database and expression levels of all matching sequences were quantified between the two conditions. As for the quantification of LINE1 and IAP repeats, the analysis was performed using the 'features' step of *SequenceImp* for repeat elements, allowing up to 3 mismatches and correcting the read counts to the number of genome mapping reads. The ping-pong signatures and 1U-10A content of the LINE1 and IAP elements were also calculated as part of this step. In all cases, normalization was performed based on the total number of reads of transcripts that remained unchanged between the two conditions. Specifically, the types of transcripts that were used for normalisation (based on the official Ensembl genebuild annotation) were: miRNAs, rRNAs, snRNAs, snoRNAs, processed\_transcripts, aa\_tRNAs, Mt\_tRNAs, other\_tRNAs, miscRNAs and RNA\_repeats. Significance was assessed using the BootstRatio algorithm<sup>21</sup>.

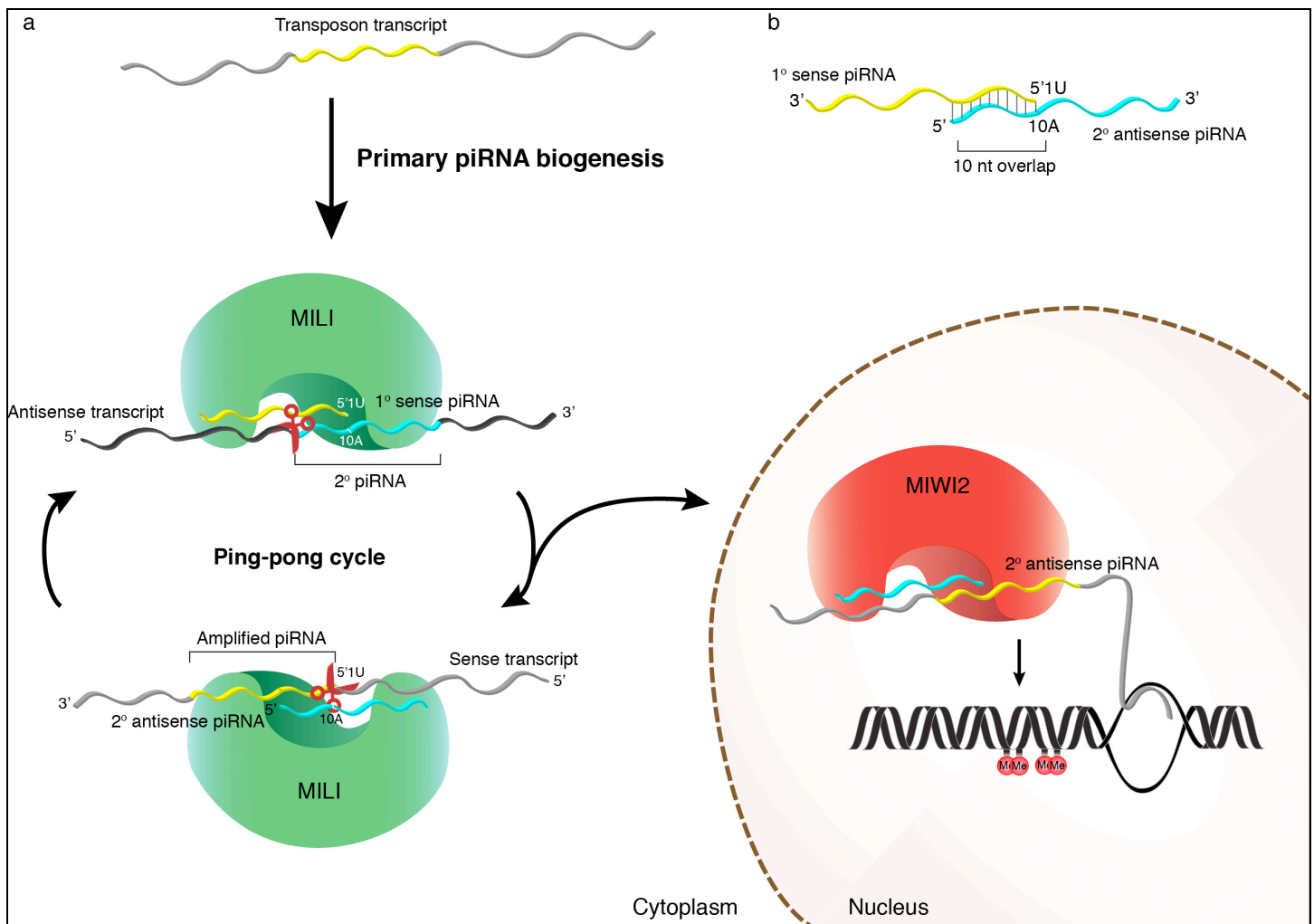
**Code availability.** Computer code used to analyze small RNA sequencing and whole genome bisulfite sequencing data is available upon request from A.J.E./D.V. and R.B./W.R., respectively.

15. Di Giacomo, M. et al. *Mol Cell* **50**, 601-8 (2013).
16. Yang, H. et al. *Cell* **154**, 1370-9 (2013).
17. Miura, F. & Ito, T. *DNA Res* **22**, 13-8 (2015).
18. Krueger, F. & Andrews, S.R. *Bioinformatics* **27**, 1571-2 (2011).
19. Nawrocki, E.P. et al. *Nucleic Acids Res* **43**, D130-7 (2015).
20. Davis, M.P., van Dongen, S., Abreu-Goodger, C., Bartonicek, N. & Enright, A.J. *Methods* **63**, 41-9 (2013).
21. Cleries, R. et al. *Comput Biol Med* **42**, 438-45 (2012).





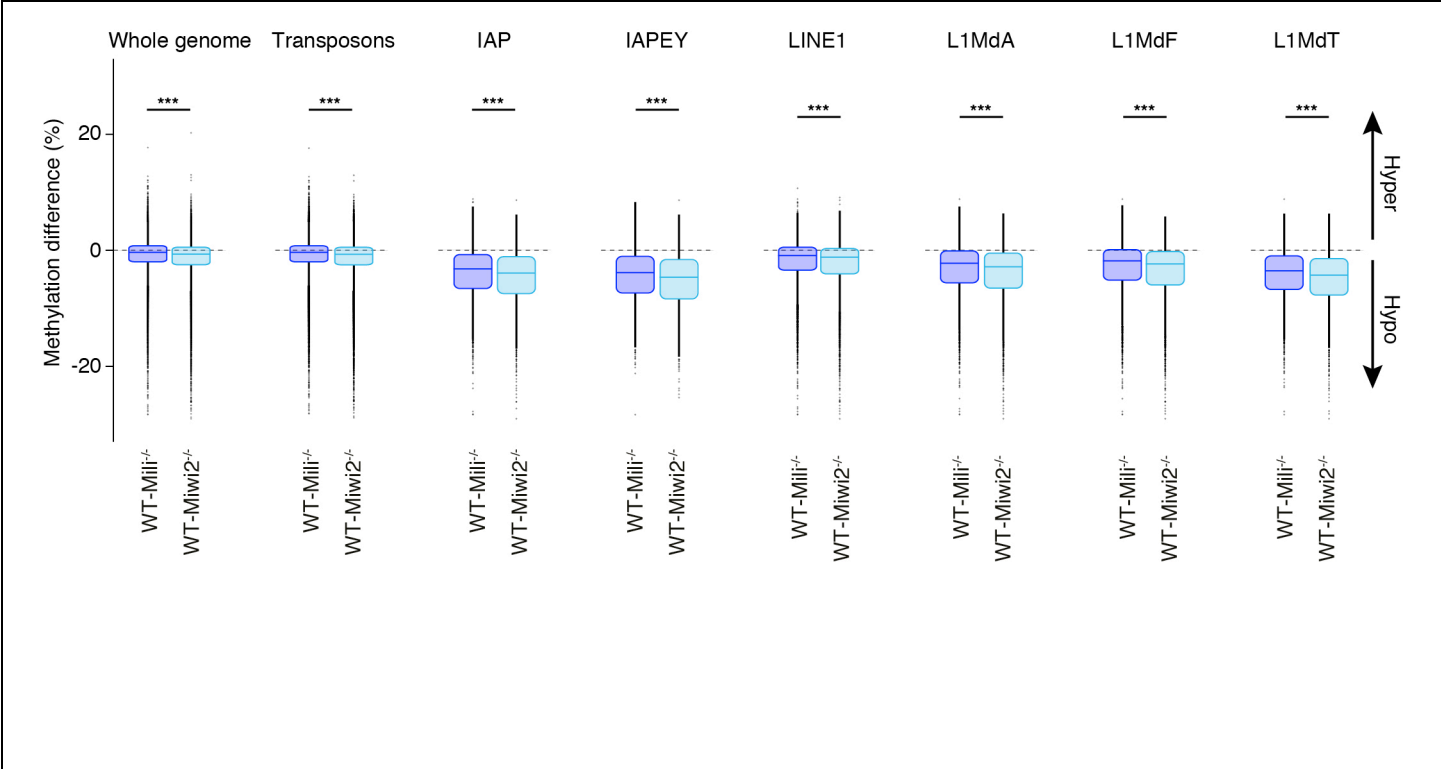




**Supplementary Figure 1**

**An overview of piRNA biogenesis during fetal mouse reprogramming.**

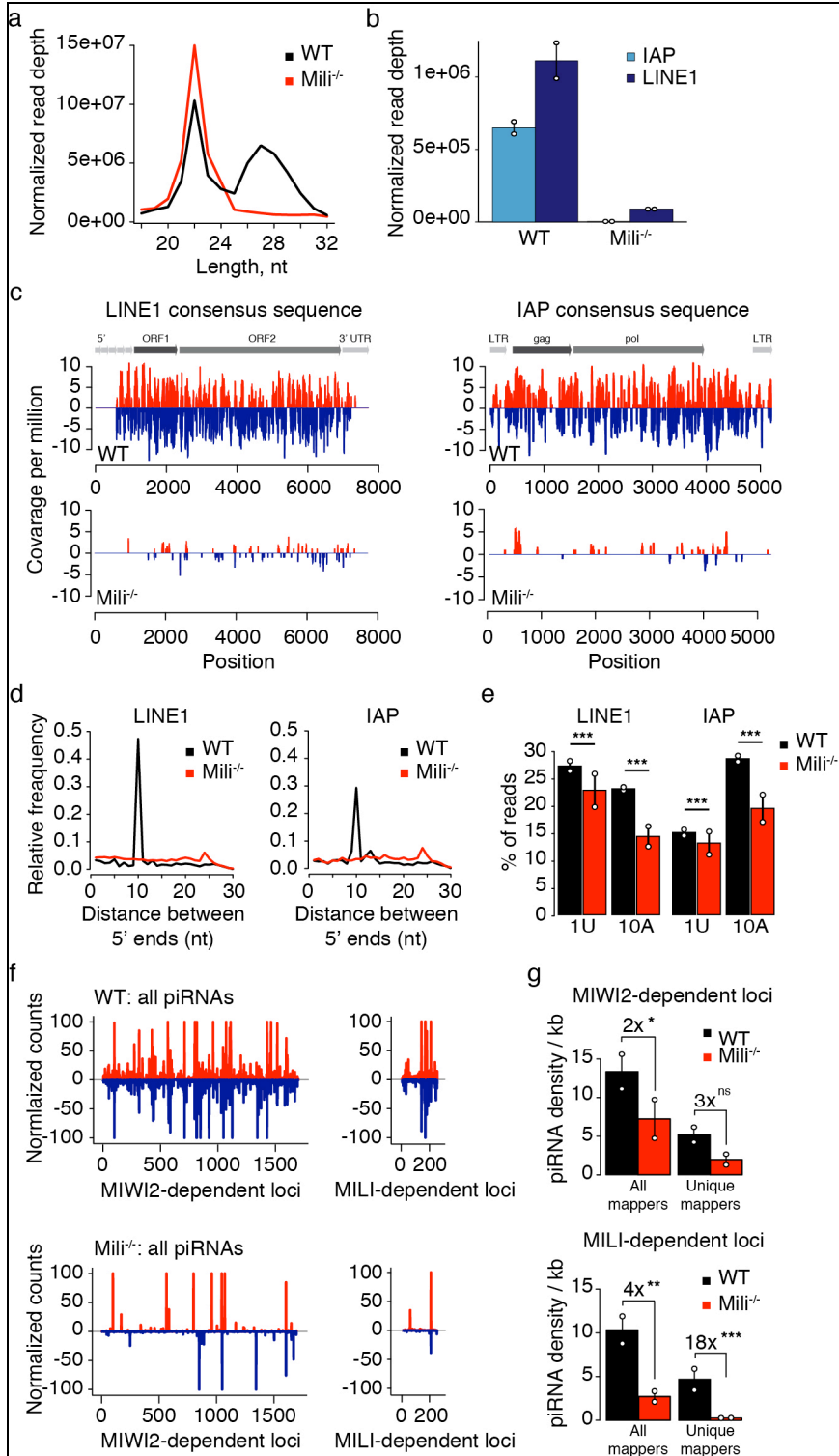
**(a)** A schematic overview of the production and amplification of a single piRNA from a transposon transcript. The transposon mRNA is processed by the primary piRNA biogenesis machinery to generate primary sense piRNAs that have a 1U bias. Indicated in yellow is a representative piRNA that will be processed from this transcript. Once bound to MILI, the piRNA guides endonucleolytic cleavage of complementary transcripts between 10th and 11th nucleotide counting from a 5' end of bound piRNA. The 3' cleavage product will be processed to form a secondary antisense piRNA, the location of which is indicated in blue. The nascent secondary piRNA is loaded into MILI or MIWI2. This licenses MIWI2 to enter the nucleus, where it is proposed to tether to a nascent transposon transcript and direct *de novo* DNA methylation by an unknown mechanism. The binding of the secondary antisense piRNA to MILI results in the piRNA-guided cleavage of transposon transcript and the biogenesis of the original piRNA. This MILI:MILI homotypic ping-pong cleavage cycle amplifies transposon-derived piRNAs during fetal reprogramming. **(b)** Features of amplified piRNA pairs. Due to the fact that most primary piRNAs exhibit 1U bias, newly generated secondary piRNAs will have a bias for A at their 10th position (10A). Complementarity between the first 10 nucleotides of amplified piRNA pairs is a key signature of the ping-pong cycle.



Supplementary Figure 2

Loss of MILI and MIWI2 has different impact on genomic and repeat element methylation.

Distribution of CpG methylation levels measured by WGBS-sequencing of *Mil1*<sup>-/-</sup> (dark blue) and *Miwi2*<sup>-/-</sup> (light blue) spermatogonia plotted as a difference to wild type (WT) spermatogonia (grey dotted line). Significance was tested using Wilcoxon rank sum test with Bonferroni multiple testing correction, \*\*\* indicates a p value of < 0.0005. An averaged value from biological triplicates (spermatogonia independently isolated from 3 animals) for each genotype is shown. The middle line represents median of the data points with whiskers indicating the data range.

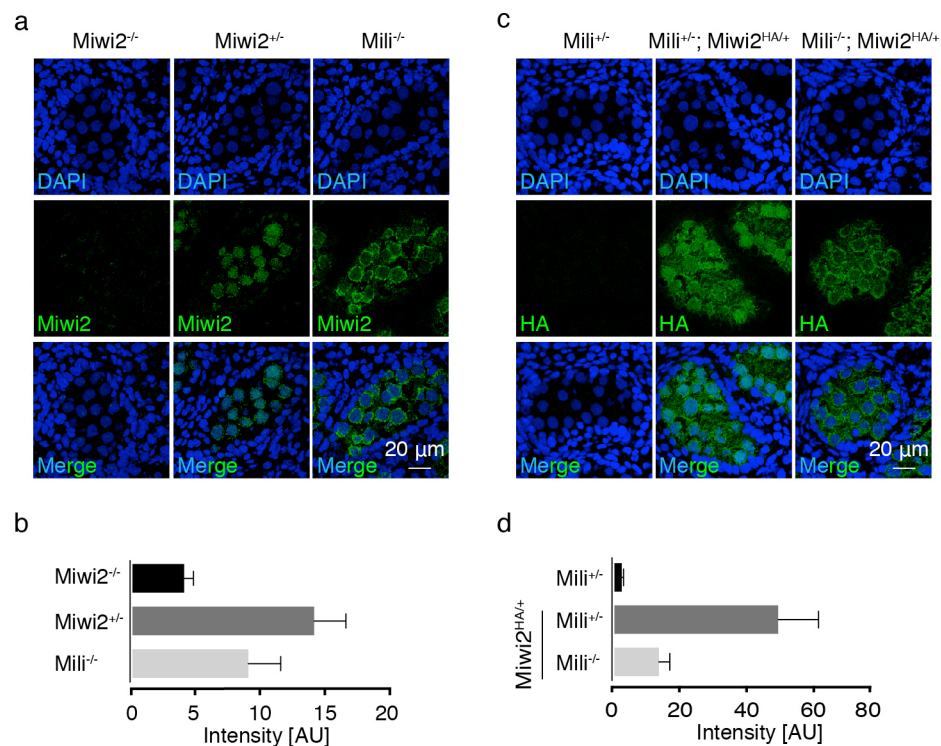


**Supplementary Figure 3**

### Transposon-related piRNAs in the absence of MILI.

**(a)** Length distribution of small RNAs in wild type (WT) and *Milir*<sup>-/-</sup> E16.5 fetal testis. An averaged value from biological duplicates is shown (n=2 animals). **(b)** Quantification of LINE1 and IAP associated piRNAs in wild type (WT) and *Milir*<sup>-/-</sup> E16.5 fetal testis. Individual

data points represent biological duplicates (n=2 animals). **(c)** piRNAs from WT and *Milr*<sup>-/-</sup> E16.5 fetal testis mapped to LINE1 and IAP consensus sequence allowing three mismatches. Positive (red) and negative (blue) values indicate sense and antisense piRNAs, respectively. An averaged value from biological duplicates is shown (n=2 animals). **(d)** Ping-pong analysis of LINE1 and IAP associated piRNAs in WT and *Milr*<sup>-/-</sup> E16.5 fetal testis. The frequency of the distance between 5' ends of complementary piRNAs for LINE1 (right) and IAP (left) is presented. An averaged value from biological duplicates is shown (n=2 animals). **(e)** Percentage of LINE1 and IAP associated piRNAs in WT and *Milr*<sup>-/-</sup> E16.5 fetal testis with a U at the first position (1U) without an A at the position 10 and an A at position 10 (10A) without a U at position 1 is shown. Individual data points represent biological duplicates (n=2 animals). **(f)** piRNAs from wild type (WT) and *Milr*<sup>-/-</sup> E16.5 fetal testis mapped to loci whose methylation is dependent upon MIWI2 and MILI, respectively. Positive (red) and negative (blue) values indicate sense and antisense piRNAs, respectively. An averaged value from biological duplicates is shown (n=2 animals). **(g)** Quantification of all or uniquely mapping piRNA densities across MILI and MIWI2-dependent loci is presented. Individual data points represent biological duplicates (n=2 animals). Significance in e and g was assessed using BootstRatio algorithm, \* indicates a p value of < 0.05, \*\* indicates a p value of < 0.005, \*\*\* indicates a p value of < 0.0005.



**Supplementary Figure 4**

**MIWI2 retains partial nuclear localization in the absence of MILI.**

**(a & c)** Immunofluorescence with anti-MIWI2 antibody (a) and anti-HA antibody (c) on fetal testis of indicated genotypes with a seminiferous tubule shown. **(b & d)** Quantification of anti-MIWI2 (b) and anti-HA (d) fluorescence signal in the nucleus of indicated genotypes. 150 gonocytes from three animals, 100 and 97 gonocytes from two animals each were analyzed in panel c from E16.5 *Miwi2*<sup>-/-</sup>, *Miwi2*<sup>+/-</sup> and *Mili*<sup>-/-</sup> fetal testes, respectively. 104, 104 and 103 gonocytes from two animals each were analyzed in panel d from E16.5 *Mili*<sup>+/-</sup>, *Mili*<sup>+/-</sup>; *Miwi2*<sup>HA/+</sup> and *Mili*<sup>-/-</sup>; *Miwi2*<sup>HA/+</sup> fetal testes, respectively. Error bars indicate standard deviation of the mean.

